

“Helicopter Theme and Variations”

Or, “Some Experimental Designs Employing Paper Helicopters”

Some possible explanatory variables are:

- Who drops the helicopter
- The length of the rotor blades
- The height from which the helicopter is dropped
- The amount of weight added to the helicopter
- Whether the helicopter is dropped indoors or outdoors
- The material from which the helicopter is made
- Which direction the rotors are bent (clockwise vs. counterclockwise)

“wilder” variations could include:

- Variations on the basic design of the helicopter
- Punching holes in the helicopter rotors

Some possible response variables are:

- How long the helicopter takes to fall
- How far the helicopter lands from a desired (point) target
- Whether the helicopter hits a desired (large) target

“wilder” variations could include:

- Whether someone is able to catch the helicopter
- How attractive the flight of the helicopter is, subjectively judged by an observer

Useful materials for helicopter experiments:

- Helicopters of several designs
- Stopwatches
- Measuring tapes or meter sticks
- Paper clips (for varying weight) or heavy binder clips for outdoor drops
- Masking tape (to mark drop heights on vertical wall and to mark target)
- Pendulum (to locate point directly beneath drop)
- Cards, dice, or random digit table
- A basket or binder clip with a long string attached will greatly ease the return of a helicopter from its landing site to the platform from which it was dropped.

Some particular experimental designs are:

One sample t :

1. “*Do we live in a vacuum?*” If air resistance has no effect on the fall of a helicopter, then the duration of its fall should be calculable from the physics law $h = \frac{1}{2} gT^2$; or, $T = \sqrt{\frac{2h}{g}}$, where $g = 9.8 \frac{m}{s^2}$, h is the height of the fall in meters, and T is the time of the fall in seconds. Have one person drop a single helicopter 30 times from a fixed height and have another person time the descents. Calculate the theoretical time to fall based on the assumption of no air resistance, and call it T_0 . (E.g. from a height of 5 meters, $T_0 = \sqrt{\frac{2 \cdot 5 \text{ m}}{9.8 \text{ m/s}^2}} \approx 1 \text{ sec.}$) Using a one-sample t test of significance, test the hypothesis $m_{time} = T_0$ vs. $m_{time} > T_0$. (Hopefully, the effect of air resistance can be detected with a very high level of significance!)
2. “*How long does this helicopter take to fall, on average?*” One person drops a single helicopter 30 times from a fixed height. Estimate the mean time for it to fall using a one-sample t confidence interval. In this case, the population of inference is only drops by this particular dropper with his or her particular helicopter. In order to generalize to a helicopter design (e.g., all long-rotor helicopters) use a random sample of 30 different long-rotor helicopters.

Note: The extrapolation to the population of all human droppers may be valid, but it could only be done based on belief, not on statistical evidence. We often do this without thinking about it. E.g., suppose that we’re testing long-rotor helicopters vs. short-rotor helicopters for some property, and for convenience, we make all the long-rotors blue and all the short-rotors yellow. How do we know that any observed differences are due to rotor length and not to color? We believe color has no effect based on plausible reasons for different helicopter behavior, but there is no *statistical* evidence that the confounding variable of color did not have an effect.

3. “*How close to a target can I get this helicopter to land, on average?*” This variation is just like experiment 2, above, except that the response variable measured is the distance that the helicopter lands from a desired point target, rather than the time it took to fall.

Paired differences t :

4. “Does this long-rotor helicopter take a different length of time to fall, on average, than this short-rotor helicopter?” Make one long-rotor and one short-rotor helicopter. All drops of both helicopters will be performed by the same person and from the same height. Flip a coin to see which one will be dropped first; after it is dropped, then drop the other one. Repeat this process 30 times, always dropping one and then the other. The paired t test of significance is appropriate for this experimental design because the helicopter drops are paired *by order*. Once the data are collected, use a t test on the paired differences to see whether there is significant evidence against $H_0 : \mathbf{m}_{\text{long-short}} = 0$ vs. $H_a : \mathbf{m}_{\text{long-short}} \neq 0$. Alternatively, create a confidence interval estimate of $\mathbf{m}_{\text{long-short}}$. In either case, the scope of inference is all the possible drops of these two particular helicopters with this particular dropper. (It is not possible, statistically, to generalize to the long- and short-rotor designs.)
5. “Do helicopters take a different length of time, on average, to fall indoors versus outdoors?” Make 25 long-rotor helicopters. Each helicopter is dropped once indoors and once outdoors from the same fixed height. As in experiment 4, do the drops in pairs, using a coin flip for each trial to determine whether the helicopter is dropped indoors first and then outdoors, or vice-versa. Measure the times of descent, and test $H_0 : \mathbf{m}_{\text{in-out}} = 0$ vs. $H_a : \mathbf{m}_{\text{in-out}} \neq 0$. Extrapolate to the population of all long-rotor helicopters dropped by this dropper. A variation would be to have a different dropper for each trial. Then the results could be extrapolated to drops by any dropper.
6. “Are Gloria and Floyd equally good, on average, at making their helicopters land near a specified target?” This experiment involves two droppers and 20 long-rotor helicopters, all dropped from the same fixed height. Choose a helicopter at random and flip a coin to see who drops it first. Both droppers drop the same helicopter, then move on to another randomly selected helicopter, again choosing who drops first by random coin flip. The response variable is the distance from a desired point target that the helicopter landed. The paired t will test the significance of evidence for $H_0 : \mathbf{m}_{\text{Gloria-Floyd}} = 0$ vs. $H_a : \mathbf{m}_{\text{Gloria-Floyd}} \neq 0$. Pairing controls for wear and tear on helicopter, increased ability over time, environmental factors, *etc.* Any significant difference may be extrapolated to Floyd’s and Gloria’s ability with all long-rotor helicopters.

Two sample t :

7. “Do short-rotor and long-rotor helicopters take different lengths of time, on average, to fall?” Construct 25 short-rotor and 25 long-rotor helicopters. One dropper (or all different droppers, for a larger scope of inference) drops all 50 helicopters from the same fixed height and their times of descent are measured and recorded. The order of the 50 drops is *completely random*—determined, say, by shuffling 25 black cards and 25 red cards and then drawing them one at a time, dropping a short-rotor for the black cards and a long-rotor for the red cards. This completely random order (as opposed to dropping the helicopters in pairs as in experiment 4) is what makes this design require two-sample t inference procedures rather than one-sample t on paired differences. Test $H_0 : \mathbf{m}_{short} = \mathbf{m}_{long}$ vs. $H_a : \mathbf{m}_{short} \neq \mathbf{m}_{long}$ or create a confidence interval estimate of $\mathbf{m}_{short} - \mathbf{m}_{long}$.
8. “Do long-rotor helicopters take a different length of time, on average, to fall indoors compared to outdoors?” 40 different long-rotor helicopters are constructed and are randomly allocated to two groups: indoor and outdoor. Each is dropped once in its appropriate location *from the same fixed height* (obviously this is crucial!) with the order of the 40 drops being completely random, as in experiment 7. The time of descent is measured and recorded for each drop. Use 2-sample t procedures to test whether the mean falling time is different for indoors and outdoors.

One sample proportion:

9. “How good is Barnaby at hitting a target with long-rotor helicopters?” 30 long-rotor helicopters are constructed. A target is drawn on the ground below a dropping platform. (The target should be large enough that it will get hit at least 10 times and will also be missed at least 10 times in order to have $n\hat{p} > 10$ and $n(1 - \hat{p}) > 10$.) One person drops all 30 helicopters in a random order and records how many of the drops hit the target. A confidence interval estimate of the proportion of hits is then constructed.

Two sample proportions:

10. “Do Zack and Xena both have the same probability of hitting a target with this particular helicopter?” One long-rotor helicopter is constructed. Each of two people will drop it 25 times towards a large target from a fixed height, with the 50 drops being performed in a completely random order (as in experiment 7). The response variable recorded is whether or not the helicopter hit the target. One may then test the hypothesis $H_0 : p_{Zack} = p_{Xena}$ vs. $H_A : p_{Zack} \neq p_{Xena}$ or create a confidence interval estimate of $p_{Zack} - p_{Xena}$.

Variations on this experiment: “Do long- and short-rotor helicopters have an equal chance of hitting a target?” “Is it easier for a person to hit a target when dropping from a lower height than a higher height?” (This latter variation would be a good example of a one-sided test.)

Chi-square goodness-of-fit test:

11. “Does this helicopter, under these environmental conditions, have a directional bias in where it lands?” Draw a “vertical” and a “horizontal” line on the ground with their intersection directly beneath the dropping point. One person drops a single long-rotor helicopter 40 times from a fixed height. The quadrant that the helicopter lands in is the response variable. Under the hypothesis of *no* directional bias, one would expect: $H_0 : p_{Q1} = p_{Q2} = p_{Q3} = p_{Q4} = 0.25$. A chi-square goodness-of-fit test will test that hypothesis. Evidence against it would suggest either a directionally-biased helicopter or, more likely, a consistent wind direction.

Chi-square test of independence:

12. “Is closeness to target (concentric circles drawn around a target on the ground like a dartboard) independent of rotor length?” Construct 30 long-rotor and 30 short-rotor helicopters. One person drops them in a completely random order towards a “dartboard” target drawn on the ground: two concentric circles separating the ground into the regions “close”, “medium”, and “far”. Count how many of each type land in each region and test the independence of rotor length and closeness to target by using a chi-square statistic on a 2×3 table. (Note: Evidence of a difference in the two does *not* necessarily indicate that either one is significantly “better”—*i.e.*, lands significantly closer on average—than the other. It only indicates a difference in the distribution of where they land.)

13. *“Is the probability of hitting a target related to drop height?”* This experiment is similar to number 12. Use “low”, “medium”, and “high” for drop heights and “hit”, “not hit” for the response variable, with counts recorded in a 3×2 table. Use a single helicopter and dropper, or else a single dropper with multiple helicopters of the same type for a larger scope of inference. A chi-square test of independence will indicate whether there is significant evidence of association between drop height and successful target-hitting.

Variation on this experiment: *“Is the probability of hitting a target related to the weight of the helicopter?”* Use paper clips, small binder clips, and larger binder clips to create different weights.

Slope of a Regression Line:

Note: both of the experiments below are best done where helicopters can be dropped from a great height (at least 5 meters).

14. *“Does adding weight to a helicopter speed up its descent?”* Construct 30 helicopters and randomly allocate them to 5 groups of 6, each group of which will correspond to a particular added weight. In a completely random order, drop each helicopter from the same height, recording the time of descent. Construct a least-squares line using the weight as the explanatory variable and the time of descent as the response variable. One may either estimate \mathbf{b} with a confidence interval or test $H_0 : \mathbf{b} = 0$ vs. $H_A : \mathbf{b} < 0$.
15. *“Do helicopters dropped from a higher height take longer on average to fall than helicopters dropped from a lower height?”* Or: *“How fast do these helicopters fall once they reach their terminal velocity?”* Construct 50 long-rotor helicopters and randomly allocate them to 5 different fixed drop heights. Then make all 50 drops in a completely random order, timing the descent of each helicopter. Construct a regression line using height as the explanatory variable and time of descent as the response variable. A test of $H_0 : \mathbf{b} = 0$ vs. $H_A : \mathbf{b} > 0$ should produce significant evidence for the latter. One may also construct a confidence interval estimate of \mathbf{b} , which is the *reciprocal* of the average falling speed of the helicopters after reaching terminal velocity. (The intercept of the regression line results from the initial period of fluttering before the helicopter “settles” into its terminal velocity. It is likely to be negative, since the helicopters tend to fall faster before they reach their terminal velocity, shortening the times of their falls.)

Experimental Design Project: Paper Helicopters

Summary of the inference procedures

“Statistical inference” refers to drawing conclusions about a population based upon data from a sample. The two types of inference we have studied are *significance tests* and *confidence intervals*. These are the contexts for which we have learned appropriate inferential procedures:

- **1-sample z.** Inference about a population mean μ when the population standard deviation σ is known. This is rarely used in practice, since it is so seldom that σ is known when μ is not.
- **2-sample z.** Comparing two population means μ_1 and μ_2 when both of their standard deviations σ_1 and σ_2 are known. This, like the 1-sample z, is rarely used.
- **1-sample t.** Inference about a population mean μ when the population standard deviation σ is *not* known and must be estimated by the sample standard deviation s .
- **Paired differences t.** Inference about the mean μ of $X_1 - X_2$, when X_1 and X_2 are paired by experimental design.
- **2-sample t.** Comparing two population means μ_1 and μ_2 when neither of their standard deviations σ_1 and σ_2 is known, and they must be estimated by sample standard deviations s_1 and s_2 . (Note: If one assumes that the standard deviations are equal, then one may “pool” the data to estimate a single s . This is seldom done in practice, since there is seldom any reason to believe that $\sigma_1 = \sigma_2$, even if we hypothesize that $\mu_1 = \mu_2$.)
- **1-sample proportion.** Inference about a population proportion p based on a sample proportion \hat{p} .
- **2-sample proportion.** Comparing two population proportions p_1 and p_2 based on sample proportions \hat{p}_1 and \hat{p}_2 . (In the case of testing the hypothesis $p_1 = p_2$, it is appropriate to estimate a single standard deviation for both \hat{p}_1 and \hat{p}_2 by pooling the data, since hypothesizing that $p_1 = p_2$ implies necessarily that the standard deviations of \hat{p}_1 and \hat{p}_2 are equal as well.)
- **Chi-square test of goodness-of-fit.** (Hypothesis test only.) Testing whether there is significant evidence against a claimed distribution of a categorical variable.
- **Chi-square test of independence.** (Hypothesis test only.) Testing whether there is significant evidence against the claimed independence of rows and columns in a two-factor table.
- **Slope of a least-squares line.** Inference about the slope b of a linear model $m_y = a + bx$ based on the least-squares regression line's slope b .

Designing experiments with helicopters

The paper helicopters you will be using this week are extremely versatile when it comes to experimental design. You can vary the height from which they are dropped, the length of the rotors, the weight you put on the bottom of them, and possibly many other things as well that you may think of.

In your group, design an experiment for six of the eight procedures described above, excluding the first two. You should briefly, but completely, describe the experiment; state what will be tested (for significance tests) or what will be estimated (for confidence intervals); and state what inference procedure is appropriate for the analysis. An example is given below.

Paired differences t. This experiment will see whether there is any difference in the average ability of two students Zack and Xena to hit a target on the ground with their helicopters dropped from the third floor of Watts.

Each student will make a helicopter, and a binder clip will be put on each one for stability. An "X" directly below the drop point will be drawn on the street outside the north Watts stairs in chalk. Each student will get three practice drops before beginning the experiment.

Then Zack and Xena will stand at the top of the stairs and flip a coin. Heads, Zack drops first; tails, Xena drops first. They each drop their helicopters, trying to hit the target. This is called a single trial. They record for each drop how far each helicopter lands from the target. Then 30 such trials are performed.

Paired differences are calculated and one-sample t inference procedures may then be used (1) to test $H_0 : \mathbf{m}_{Z-X} = 0$ vs. $H_A : \mathbf{m}_{Z-X} \neq 0$ and (2) to estimate \mathbf{m}_{Z-X} . (Note: the inference can only be drawn about Zack with his helicopter and Xena with hers. This experiment does not determine whether any difference is due to the person or the helicopter, since these are confounded.)

Write up each of your group's six experiments neatly on the blank paper provided; this is to be handed in. Be prepared to present any of them to the class and answer questions people may have about them.

Make a note of those that you would most like to perform. Each group will execute some, but not all, of their experiments.